**Unless it is specified, you may use Word, Excel, Power Point, Python and R to answer the questions in this exam. There are a total of 4 (four) multi-part questions, with point values noted for each question.**

**Please show your calculations, or the details of your program(s), for each problem. The program(s) should be commented so that each step is clearly explained.**

**Combine all of your answers/files into a single zipped file and post the zipped file.**

**Problem #1: (40 points)**

**Cluster NY zip codes into 4 clusters using the NY\_ZIP.csv dataset in CANVAS, R/Python, and features** **Returns\_pct1 to Returns\_pct6. Compare the four clusters for each of the following two methods.**

* **Hierarchical clustering (centroid)**
* **K-means, population (centroid)**

**Problem #2: (20 points)**

* **Load the Titanic dataset from CANVAS**
* **Convert the categorical variables to 0,1 indicators**
* **Store every fourth record in a “test” dataset starting with the first record**
* **Store the rest in the “training” dataset**
* **Use ANN with 6 hidden nodes to classify passengers (survival=1 vs. 0).**
* **Measure the performance of the model against the test data.**

**Problem #3: (20 points)**

* **Load the Titanic dataset from CANVAS**
* **Store every fourth record in a “test” dataset starting with the first record**
* **Store the rest in the “training” dataset**
* **Use CART to classify passengers**
* **Measure the performance of the model against the test data.**

**Problem #4: (20 points)**

* **Load the Titanic dataset from CANVAS**
* **Store every fourth record in a “test” dataset starting with the first record**
* **Store the rest in the “training” dataset**
* **Use C5.0 to classify passengers**
* **Measure the performance of the model against the test data.**